A Grid Enabled E-Theses and Dissertations Repository System

Por Yee, Ong Ying, Delina Yin, and Maizatul Ismail
Faculty of Computer Science and Information Technology, University of Malaya, Malaysia

Abstract: *Some of the universities in Malaysia are still implementing Hybrid Electronic Theses and Dissertations (ETD) approach in managing Theses and Dissertations (TD). One of the limitations of the Hybrid ETD approach is its online cataloguing method, which is only available at the physical location of the TD instead of enabling the information to be retrieved online. Maintaining the performance and the data accessing rate of an ETD system has become challenging, due in part to the high number of scholars who utilise and access the system. In order to allow remote access and maintain the services (such as scalability, accessibility, availability and expressibility), a Grid Enabled E-Theses and dissertations repository system (GREET) has been proposed in this paper to provide uniform access of knowledge integration among distributed heterogeneous platforms and repositories by using data grid technology. Comparative performance results between a non-grid architecture and GREET has been benchmarked. It has been proven that GREET is able to increase the processing time approximately three times faster than the non-grid architecture. Furthermore, multiple file streams can be opened to support larger volume and larger capacity of file operation so that GREET is able to decrease the chances of network congestion caused by input/output file operations. For future direction, research will be focused on searching algorithm using data mining or pattern discovery to minimise the respond time.*

1. Introduction

Electronic Theses and Dissertations (ETD) is an electronic repository system which comprises information such as concepts and outputs of the latest researches produced by scholars during their process of cognition, exploration and analysis. In order to improve the ETD system in Malaysia, a literature review especially on the evolution of the ETD system in Malaysia and other relevant countries have been studied, compared and analysed. From the analysis result, we have identified that a good ETD system must be able to provide the following services (Please refer to Figure 1 for the proposed ETD repository services):
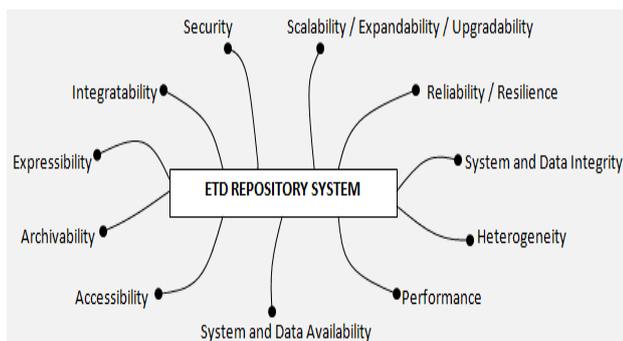

Figure 1. Proposed ETD repository services.

1. Scalability/Expandability/Upgradability: referring to the ability of an ETD system to expand without having significant degradation in performance.
2. Reliability/Resilience (Fault Tolerance): referring to the ability of an ETD system to handle any unexpected condition and provide disaster recovery facility when needed.
3. System and Data Integrity: referring to the ability of an ETD system to ensure the consistency and validity of the information stored after a particular processing such as data duplication, data insertion, deletion and modification.
4. Heterogeneity: referring to the ability of an ETD system to handle multiplicity of resources that are heterogeneous in nature and will encompass a vast range of technologies [16].
5. Performance: refers to the turn-around time for a complete uploading or downloading processes of an ETD system.
6. System and Data Availability: refers to the uptime of an ETD system.
7. Accessibility: referring to the availability of the information throughout the searching and retrieving processes via an ETD system.
8. Archivability: referring to the ability of an ETD system to catalogue and store divergent data type.
9. Expressibility: referring to the presentation of the information that an ETD system can provide.

10. Integratability: referring to the ability of an ETD system to integrate with other existing ETD systems.

9. Security: Referring to the ability of an ETD system to avoid malicious and other security threats from hackers and crackers.

From the literature, we have found out that most of the universities in Malaysia, in particular public universities funded by the Malaysian government are still implementing the Hybrid ETD approach in managing theses and dissertations (the definition of Hybrid ETD can be obtained in the following section). One of the drawbacks of implementing the Hybrid ETD approach is due to its nature of online cataloguing method which can only be used to pin point the physical location of the theses and dissertations instead of making the information retrievable online. Due to this issue, a lot of scholars in Malaysia are still suffering as information retrieval still needs to be done manually.

Apart from the availability issue, the quality of the information retrieved has also become an important factor when an ETD has the facility to reveal the same information to the scholars in a more presentable and meaningful manner. Retrieved ETD system data can portray contents using colour diagrams and images, as well as dynamic elements such as animation and flash presentation, and multimedia resources such as video and audio. This is in contrast to printed material, which some scholars may find boring.

However, the implementation, management and integration of heterogeneous operating system, divergent data type and storage server is a challenge to the Malaysian ETD systems. At present, there is no uniform search browser which enables scholars to search for any related information which is obtained from different ETD system at Malaysia although there are some ETD systems that are maturely integrated, for instance *myais, mymanuskrip*[5]*, eprints* and etc.

Table 1. Student Intake for Public Institutions of Higher Learning (IPTA) in Malaysia from 2002-2007 [17].

Year	2002	2003	2004	2005	2006	2007
Student Intake	64,061	70,481	81,075	80,885	89,633	128,839

Scalability is another main factor to ensure the performance and durability of an ETD system used. Based on the survey result shows in the Table 1, the volume of the number of thesis and dissertation that will be produced by the scholars will increase due to the rapid growth in the yearly intake [17]. Thus, an assumption has been made that the number of scholars who will be using an ETD system to search of a research publication will also increase. As a result, the high demand of data transaction, maintaining the performance and the data accessing rate of an ETD system in future will be an impact and challenge especially for the ETD systems in Malaysia. Moreover, integrating the current existing ETD system with other additional ETD systems from time to time will become tedious to most of the ETD system developers in Malaysia. Therefore, in order to maintain the services and alleviate the problems which have been mentioned earlier, a Grid Enabled E-Theses and Dissertations Repository System (GREET) has been proposed.

The remaining of the paper is divided into three sections. In the Section 2, we reviewed and evaluated the growth of the ETD systems as in general. Then, Section 3 depicts about the related work which consists of i) comparison between the growth of the ETD systems in Malaysia and other related countries and ii) evaluation and analysis of current existing ETD systems in Malaysia have been carried out. The reason of implementing the proposed system together with its preliminary result will be discussed in Section 5 after revealing the GREET architecture in Section 4. In Section 6, we draw some conclusions and discuss about the future work.

2. Growth of the ETD System

From the literature, the evolution of TD repository systems can be classified into three stages: Manual TD, Hybrid ETD and ETD. Paper-printed format which are collected during the olden days are categorised under Manual TD. Besides, Manual TD also includes collections of theses and dissertations in book and microfilm (a picture capturer created by George McCarthy to take pictures of document in black and white photography [3]). At Manual TD stage, the technology of microfilm has been widely used to capture series of documents relating to students' theses and dissertations. However, the tape-liked microfilm can only be viewed using a microfilm machine. Due to this reason, the above-mentioned method has encountered several disadvantages especially in scalability/expandability, accessibility, availability and expressibility.

As mentioned in the previous section, the collection of theses and dissertations is increasing every year. As a result, this phenomenon has enforced the scalability/expandability of infrastructures used in which more money and time are needed to extend the physical storage (building) for storing theses and dissertations.

In terms of accessibility, scholars are facing difficulty in obtaining resources due to geographical barriers. Moreover, it is very costly and time consuming for a scholar to travel from one place to another just to obtain the required resources. Besides, most of the libraries have limited copies of paper printed theses and dissertations. If a scholar has borrowed or occupied a copy of the resources, other

scholars will not have an opportunity to obtain that copy of resources.

In terms of expressibility, most of the paper printed TD are not expressive enough due to its presentation manner. Therefore, scholars or readers might lose interest in reading printed materials which are in typewriting or black and white microfilm format. Furthermore, there is a high impact for the resources to be recovered if the paper printed copy of TD is destroyed or damaged.

The emphasis of knowledge sharing by digitalising theses and dissertation has been mentioned by Davinson in 1977 [6]. The National Digital Library of Theses and Dissertations (NDLTD) [20] supported his view, and subsequently conceptualises the full model of online theses and dissertations in 1987. Electronic Theses and Dissertations (ETD) research field is pioneered by Virginia Tech, US in NDLTD project. The NDLTD project is initiated by Edward Fox which is the current director of NDLTD and professor in Virginia Tech [8]. However, the initial development of NDLTD is on cataloguing the paper-based and microfilm theses and dissertations. The online TD cataloguing is a Hybrid ETD approach which combines the computer technology usage to enable a user or scholar to search for the physical location of a thesis or dissertation. According to [28], the hybrid ETD approach is able to reduce the duration of the searching time and increase the accuracy of identifying a resource. Subsequently, the Hybrid ETD approach may partially solve the accessibility problem but the knowledge sharing is still restricted by geographical barriers. When considering the actual location of obtaining the resources, scholars are still required to travel in order to obtain the resources. As a result, the Hybrid ETD approach is still unable to solve the scalability, accessibility, availability and expressibility issues.

In 1987, University Microfilms, Inc (UMI) integrates Standard Generalized Markup Language (SGML) into Hybrid TD. According to [7], the use of SGML in Hybrid TD allows far more complex searching. For instance, for fully marked up documents, searches can be made on bibliographic citations or such citations can be extracted from each TD to create a citation database as a secondary product. SGML is also an independent platform, in a way that a single document can be shown successfully on any number of computers without conversion. However, SGML lacks the proprietary coding which can cause word-processed documents difficult to transfer between applications and platforms. Besides, the integration process is not cost effective due to the immature level of Information Technology (IT) at that time. As a result, in the early of 1990s, NDLTD has begun to seek for global collaborations among the institute from different countries and changed its name from National Digital Library of Theses and Dissertations to Networked Digital Library of Theses and Dissertations (NDLTD) [9, 20]. Apart from the

global collaborations, more efforts have been done to shorten the distance of knowledge sharing virtually via web technology. However, there are still limitations in the hybrid TD where resources that can be searched and retrieved are only limited to certain metadata such as title and abstract. Consequently, the details of a TD still need to be obtained from a specific location.

In term of scalability, ETD is able to scale by increasing the secondary storage such as a data server or increasing the hard disk spaces. Moreover, maintaining and safekeeping the TD in electronic format can also be done by using the secondary storage. As a result, ETD is less costly if compared with the Manual TD and Hybrid ETD stages which require large physical storage (building) for storing. Besides, ETD is able to duplicate the TD as a backup to avoid any information loss during storage failure.

The actual implementation/realisation of NDLTD has been achieved in 1997 by Virginia Tech. Besides, Virginia Tech has become the first university which implemented a policy about the electronic submission of students' theses and dissertations. Following in the lead set by Virginia Tech, many universities have started to implement a similar approach by collecting electronic theses and dissertations from students. For example: Universities in Korea in late 1990s [15] and University of Latvia [14]. By implementing ETD, TD not only can be searched by using online catalogue but also downloadable via web browsers. As a result, scholars no longer face availability and accessibility problems which happen in Manual TD and hybrid ETD stages.

In terms of expressibility, students are allowed to enhance and enrich the contents of their TD by adding multimedia elements and colour images. Scholars who search or obtain the resources will find it more beneficial compared to the paper-printed based method. For instance, animated image or video can be used to explain an operation in order to increase understanding power of scholars rather than obtaining information from a plain image.

According to [23, 20], NDLTD is now researching on how to link the TD from outside source such as the NSDL (National Science Digital Library) through XML or union catalogue (metadata catalogue) to enable universal accessibility of ETD. Table 2 shows the service-oriented analysis and synthesis among Manual TD, hybrid ETD and ETD.

Table 2. Service-Oriented Analysis and Synthesis among Manual TD, Hybrid ETD and ETD.

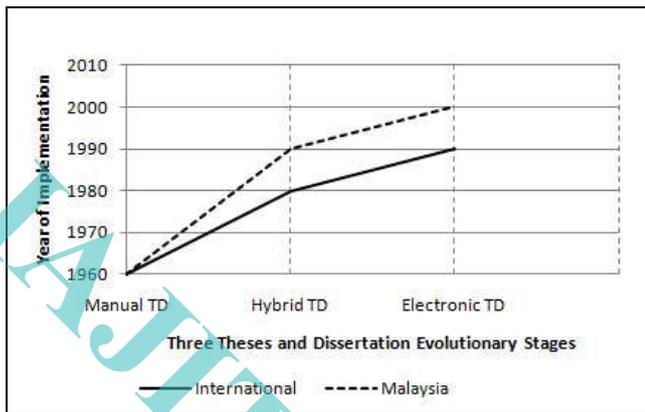| Services	Manual TD	Hybrid ETD	ETD
Scalability / Expandability / Upgradability	Costly (in terms of expanding physical storage) Trade-off: Searching time for TD will be longer once physical storage is expanded	Costly (in terms of expanding physical storage)	Less cost (can be expanded using computer storage such as a data server or expand the hard disk spaces.)
Reliability / Resilience	Low reliability (no backup nor duplication of TD)	Low reliability (no backup nor duplication of TD)	High reliability (backup and duplication can be done by using secondary storage.)
System and Data Integrity	N/A	Steps and algorithms need to be integrated into system to ensure the data and system integrity with users right control.	Higher data integrity and system integrity need to be guarded because ETD enables full access to the theses and dissertations online.
Heterogeneity	N/A	Able to support heterogeneous platforms and database management systems.	Able to support heterogeneous platforms, physical storage, databases and database management systems.
Performance	Depend on the location of the TD stored. Longer time is needed to identify the actual location of the TD stored Time needed to search for a TD at a particular library	Depend on the location of the TD stored. Actual location of the TD stored can be identified via the Internet. Time needed to search for a TD at a particular library	Depend on the downloading time. Whole TD can be downloaded from the Internet.
System and Data Availability	TD can be obtained at libraries Access time is depend on the library working hours	Searching and retrieving resources only limited to certain metadata such as title and abstract. TD still need to be obtained at libraries Access time is depend on the library working hours	Whole TD can be obtained online Access time is 24 hours
Accessibility	Low accessibility (data need to be accessed manually)	Low accessibility (data need to be accessed manually)	High accessibility (data can be retrieved via a web browser)
Archivability	Catalogue in manual format Data stored in physical location	Catalogue in electrical format Data stored in physical location	Catalogue in electrical format Data stored in electrical storage (e.g.: a data server)
Expressibility	Low expressibility Paper-based and microfilm Black and white printout	Low expressibility Paper-based and microfilm Black and white printout	High expressibility Consists of multimedia elements such as animated images colour printout
Integratability	N/A	Only same catalogue server can be integrated	Both catalogue server and storage server need to be integrated
Security	N/A	N/A	Basic security services such as user authentication, user access control and backup server can be implemented.

3. Related work

Figure 2. Evolution of TD Repository System in Malaysia and Other Related Countries.

A comparison of the growth for the three stages (Manual TD, Hybrid ETD and ETD) in Malaysia and other relevant countries is illustrated in Figure 2. As in the other related countries, Malaysia has started the Manual TD stage since 1960s [11]. According to [11], University of Malaya (one of the universities in Malaysia) has enforced postgraduate students to submit their theses and dissertations in hard cover or book format since 1960s. However, from the graph, the development of TD repository system in Malaysia is much slower compared to the other related countries. According to [4], Online Public Access Catalogue (OPAC), which is the first Hybrid TD, was only implemented by University of Malaya in 1992 for easing scholars in searching resources.

In 2005, PERPUN (Conferences of University Library and National Library of Malaysia) initiates MyTO (Malaysian Theses Online) [19] that aims on listing theses collections compiled from participated institutions. MyTO is the collaboration efforts cooperated by National Library of Malaysia and twenty-three universities in Malaysia which include twenty public universities and three selected private universities. There are approximately 23,000 TD in MyTO contributed by twenty-one universities [21]. As similar to other online cataloguing systems, MyTO can only be used to identify the physical location of a resource. Therefore, MyTO does encounter the scalability, accessibility, availability and expressibility problems which have been mentioned in the earlier section.

In 2007, University of Malaysia Perlis (UniMAP) has deployed DSpace in building a digital repository named iRepository [26] (one of the earliest ETD systems in Malaysia). iRepository enables scholars to access and download TD, conference and journal publications and other communities' related information which associates to UniMAP via web browsers.

DSpace is an open source software which aims to provide easy but comprehensive services to users, authors and librarians to build their digital repositories. DSpace is a widely used software system for digital library which has been conceived by Hewlett-Packard (HP) Labs and developed in conjunction with Massachusetts Institute of Technology (MIT) library. However, according to [24, 29], DSpace can only be used in UNIX system and it provides a generic end-user interface design which is hard to be tailored by typical end users.

On the other hand, University of Technology MARA uses Alchemy Web Server as a digital document repository for populating abstract of students' TD in 2001 [25]. According to [13], Alchemy Web Server enables web browser-based query, retrieval, and viewing of resources stored in Alchemy database. Moreover, Alchemy Web Server supports full text and metadata searching. However, Alchemy Web Server has interoperability issue where it can be installed and executed only in Windows-based platform. Besides, Alchemy Web Server is a business application which provides three pane contextual views on its interface and cannot be tailored by users.

According to [12], University of Malaya (UM) has deployed a different tool named Greenstone Digital Library Software to develop the digital repository for handling Malay Manuscripts [5]. Greenstone is an open source digital library software suite developed by New Zealand Digital Library Project at the University of Waikato [10]. The features provided in Greenstone are aimed for allowing users to assemble their own digital libraries of online documents and then provide the means to disseminate them over the web or by Compact Disc (CD). The advantage of using Greenstone is that it is able to be executed on heterogeneous operating systems (OS) such as Windows, UNIX and Mac OS platforms. Moreover, for users who have the knowledge in programming, they can personalise their user interface together with using any suitable plug-in at the client side such as Database plug-in, Metadata plug-in, Word plug-in and so forth. However, Greenstone only allows user with an administrator role to manipulate the system. For instance, an administrator has to create collections, define system's parameters and identify authorised users so that they are able to add or personalise their collections. As a result, Greenstone is not advisable to be used for institutional repository or any integration that consume huge volume of resources.

In Malaysia, the concept of knowledge sharing is still not fully implemented among universities in Malaysia. Although theses and dissertations are collected in digital format, but the expandability of the technology used is still limited. In order to alleviate the ETD development issues which have been mentioned earlier and to increase and challenge the advancement of ETD towards better level of achievement, a

prototype which uses the integration of ETD and data grid technology has been proposed and developed.

4. System Design and Implementation of GREET

Data Grid Electronic Theses and Dissertations (GREET) is proposed to provide uniform access of knowledge integration among distributed heterogeneous platforms and repositories by using data grid technology. The implementation of GREET utilises multi-tiered client-server architecture as shown in Figure 3. GREET comprises of three main tiers which includes Presentation Tier, Application Tier and Data Tier. The Presentation Tier is the outer layer in GREET which directly communicate with the end user (client). The main function of the Presentation Tier is to provide an interface to the communications environment which is used by the application processes such as translating tasks and results (unformatted response and raw data) into the format which is understandable and readable by clients.
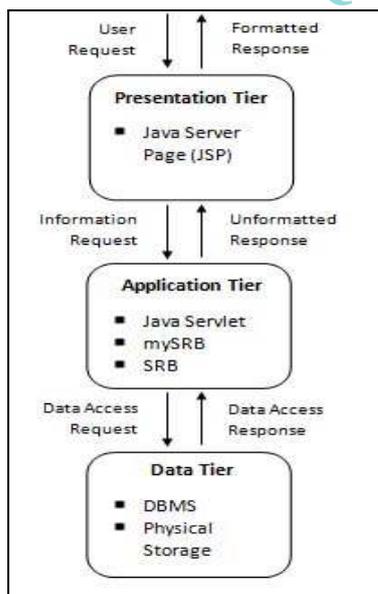

Figure 3. The Architecture and interaction cycle of GREET.

Figure 4 shows an instance of a default GUI of mySRB. mySRB is a web-client provided to support the data accessing and metadata brokered by SRB [18, 22, 27]. However, in the development of GREET, mySRB is not adopted entirely as a Presentation Tier application due to the fact that mySRB does not provide customisation of interfaces based on user authority which may raise the copyright issues and insecurity of information in the ETD system. For example, clients from the same federation are able to access and modify any data in the federation although they are having different user authority due to mySRB is using domain name rather than user's rights for data accessing.
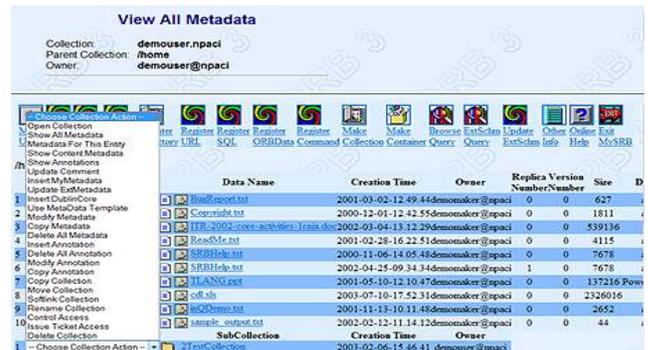

Figure 4. Default GUI of mySRB.

Besides, another issue which draws our concern towards the implementation of mySRB is the usability issue. mySRB uses a compact single web interface to display all the functionalities in a long-listed drop down menu [21]. Besides, the same web icon, 🌀, has been used repeatedly to represent different functionalities. As a result, tedious searching process and misidentification of a proper function used may occur among mySRB users.

In order to alleviate the issue mentioned above, an enhanced Presentation Tier web application which uses Java Server Pages (JSP) has been developed and added on top of mySRB. Figure 5 shows the enhanced Presentation Tier web application prototype used in GREET.
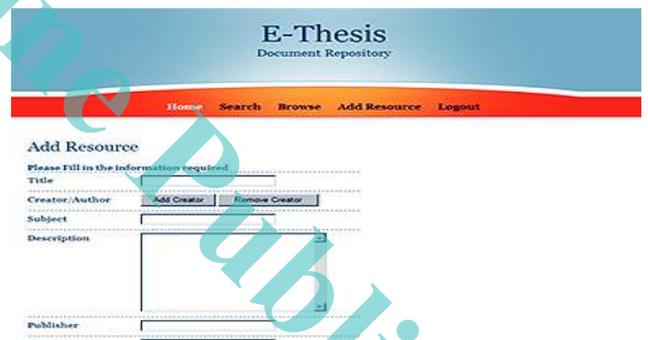

Figure 5. Enhanced Presentation Tier Web Application Prototype of GREET.

As shown in Figure 5, the enhanced web application is simple, easy to use and yet is able to provide compact functionalities to the clients. In addition, the enhanced web application allows the appearance to be customised according to the user authority such as administrator, registered client and unregistered client. Administrator is able to grant access rights such as retrieve, upload, delete and modify data within a federation. For unregistered clients, they are allowed to view the TD information such as title, date, writer and abstract of the TD whereas for registered clients, they are able to grant an extra permission to download and view a particular TD. As a result, the data integrity issue can be solved due to the fact unauthorised users can no longer modify or delete any data within the

federation after implementing the enhanced web application.

The Application Tier contains logic to coordinate the GREET and process requests which are sent by clients. Every request sent by a client is pulled into the Application Tier. Then, the request is calculated and processed according to the embedded algorithm and logic to produce logical decisions and evaluations. Once a request has been issued by a client, the JSP in the Presentation Tier calls the service () method in the servlet to serve the request (Please refer label b in Figure 6). Then, the service () method identifies appropriate methods to handle the request based on the request parameters (Please refer label c in Figure 6). The JSP calls the init () method to set the servlet into the ready mode before the servlet is able to handle the following request (Please refer label a in Figure 6). When the request submitted by the client is involved with information access through federation or Data Tier, the request is passed to mySRB and SRB Server for further processing.
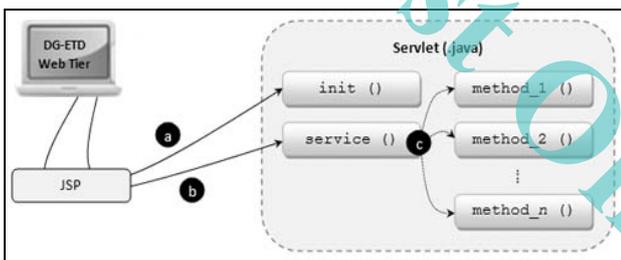

Figure 6. The Java Servlet Request Interaction.

Data Tier is the location where data is stored and retrieved from a database or a file system. Initially, the SRB Server searches data physical attribute in MCAT (Metadata Catalogue) when mySRB redirects the request to the SRB Server. MCAT is one of the components in SRB which is responsible to keep track of logical name spaces and metadata mappings (unformatted structure of host name and file path which correspond to the physical location) before sending back to the SRB Server. Once the SRB Server gets the data from the physical repository, the data is then returned to the requestor (servlet). Consequently, servlet sends the data to JSP and displays to the client in a formatted view (please refer to Figure 7 for the mySRB, SRB Server and MCAT interaction cycles).
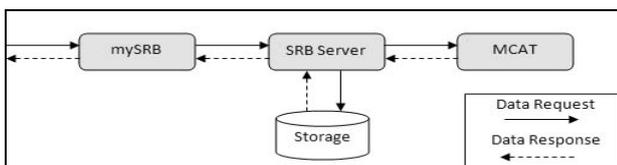

Figure 7. mySRB, SRB Server and MCAT Interaction Cycles.

In order to enhance the performance and scalability especially on integration with other ETD systems, GREET has integrated the federated MCAT SRB approach in the deployment architecture. Federated MCAT SRB is the new released version of SRB. The federated model allows two or more standalone SRB systems to interact with each other. Besides, it also provides seamless access of data and metadata across SRB systems. Moreover, federations can be added by administrator using Java Admin Tools or using Szone command in Scommands. For SRB version 3.3 and above, federations can be easily ingested by using a new script named zoneingest.pl provided along with the SRB installation script. Figure 8 shows the overview of the federated model of GREET when integrating with multiple federations.
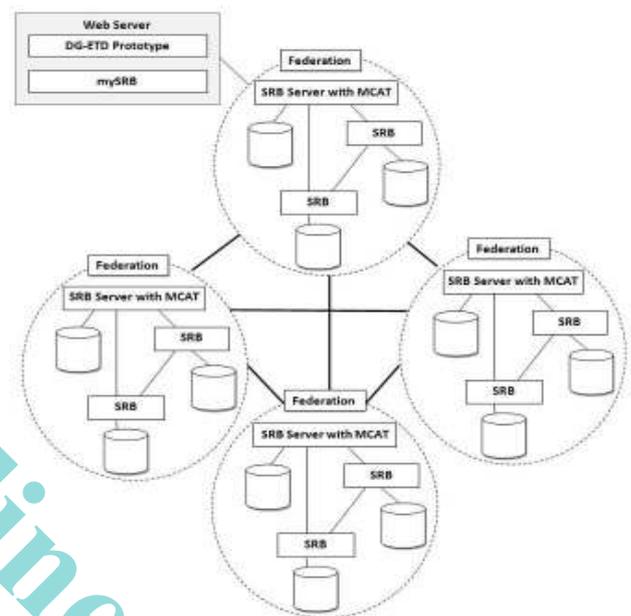

Figure 8. Federated GREET deployment Architecture.

To illustrate the schematic view of a federated environment between mySRB and SRB in the GREET deployment architecture, a scenario of two federations and four SRB Servers has been developed and shown in Figure 9. When a client is requesting a data named file_z which located in Storage B, the request is being sent to the JSP and servlet in the Presentation and Application Tiers respectively. Then, the request is sent to the mySRB and redirected to the SRB Server A. The SRB Server A then sends the request to the SRB Server B. The MCAT at the SRB Server B searches for the required data. If the required data cannot be found in the Federation A, the request is being passed to the SRB Server D. The MCAT at SRB Server D does the same process as before for searching the data in Federation B. When the data can be found at the Storage B, the SRB Server D redirects the request to the SRB Server C which contains the data. Subsequently, the SRB Server C retrieves the file_z from the Storage B and passes the file_z to the SRB Server A directly. Once the SRB Server A has obtained the requested data, the file_z is being sent to the

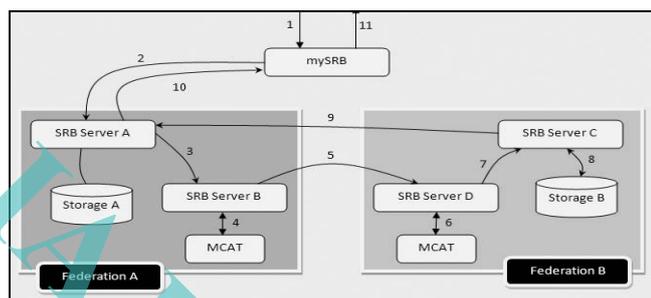requestor through mySRB, servlet and finally to the JSP.

Figure 9. Schematic of a Federated Environment between mySRB and SRB.

From the scenario above, it shows that by implementing federated MCAT SRB approach in the GREET deployment architecture; the scalability of ETD systems can be achieved since federations is able to be easily modified, added, or removed to accommodate the load scalability. Besides, the performance of GREET can be enhanced after applying the federated MCAT SRB approach because it enables load balancing among MCATs.
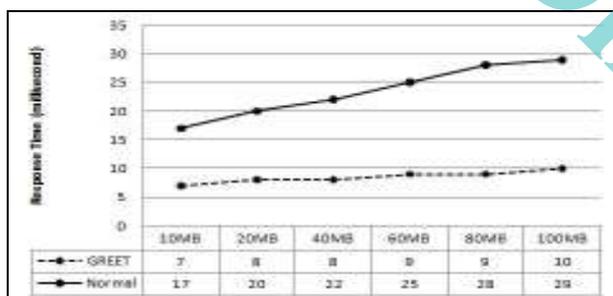
5. Results and Discussion

Figure 10. Performance Comparison among Non-Grid and GREET Architectures.

Figure 10 shows the performance results by benchmarking with non-grid architecture (normal architecture). The following size of the files: 10 Megabytes (MB), 20MB, 40MB, 60MB, 80MB and 100MB are prepared and used as the test materials for benchmarking in the performance testing. AppPerfect Test Studio has been used to evaluate the performance of the aforementioned testing [1]. All benchmarking testings are run in the same machine (Pentium Core 2 Duo Centrino 2.0 GHz and 2GB memory) connected via network to the application server and storage server. Both upload operations are recorded and executed for 30 attempts before an average response time has been obtained.

From the testing, when the capacity of a file size increases, the normal architecture requires longer time to complete the task because it has only one file stream for job submission and execution. However, GREET is able to speed up the processing time to approximately three times faster than the normal architecture (please refer to Figure 11 for the normal architecture and GREET architecture file uploading operations).
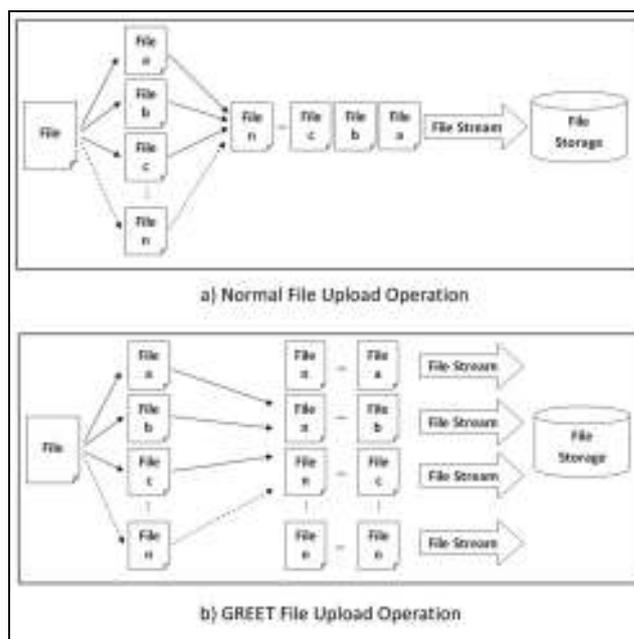

Figure 11. A Comparison of Upload Operation between Normal File and GREET.

When an input/output (IO) operation on heavy file is carried out using normal architecture, network congestion may happen when the file is split into multiple parts that needed to queue for an operation.

6. Conclusions and Future Work

In this paper, a Grid Enabled E-Theses and Dissertations Repository System (GREET) has been proposed. The enhanced presentation tier web application in GREET is able to alleviate usability issue and customise interfaces based on user authority.

By using GREET, multiple file streams can be opened to support larger volume and larger capacity of file operation so that GREET is able to decrease the chances of network congestion caused by file operation. As a consequence, GREET indirectly helps to improve the reliability and availability of the IO operation.

For future enhancement, more research will be carried out and focused on searching algorithm using data mining or pattern discovery. Searching performance in terms of response time and relevance of data is crucial especially when volumes of data become larger. Subsequently, research on finding relevant data via different storage server can be carried out to improve the response time in GREET. Besides, the architecture of GREET can be designed to be able to integrate with other existing storages, web applications and even web services. Reengineering web applications to web services proposed by [2] can be revised and used as a stepping stone for the future development of GREET.

Acknowledgements

References

[1] AppPerfect Official Website, 2003 [Online]. Available: http://www.appperfect.com/ (accessed 6 April 2009).

[2] Bouchiha D., Malki M. and Mostefai A. E. K., "Towards Reengineering Web Applications to Web Services", *The International Arab Journal of Information Technology*, vol. 6, no. 4, pp. 359-364, 2009.

[3] Cady S. A., "Microfilm Technology and Information Systems", in Bowden M. E., Hahn T. B., and Williams R. V. (Eds.), *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems*, pp. 177-186, 1999.

[4] Chan S. N., "Making Information Literacy a Compulsory Subject for Undergraduates: The Experience of the University of Malaya", *World Library and Information Congress (IFLA Journal)*, vol. 21, no. 29, pp. 328-335, 2003.

[5] Digital Library of Malay Manuscript (MyManuskrip), 2008 [Online]. Available at: http://mymanuskrip.fsktm.um.edu.my/Greenstone/cgi-bin/library.exe (accessed 3 February 2009).

[6] Davinson D.E., "Theses and Dissertations as Information Sources", C. Bingley, Linnet Books, London, Hamden, 1977.

[7] Dobratz S., Schulz M., Potter P., "SGML/XML-based Electronic Theses and Dissertations: Existing Projects and Standards", *The Internet and Higher Education*, vol. 4, no. 2, pp. 93-104, 2001.

[8] Fox E. A., "Overview of a Guide for Electronic Theses and Dissertations", *Digital Libraries and Virtual Workplaces: Important initiatives for Latin-America in the Information Age*, Washington, Organization of American States, pp. 125-156, 2002

[9] Fox E., Eaton J., McMillan G., Kipp N., Weiss L., Arce E. and Guyer S., "National Digital Library of Theses and Dissertations: A Scalable and Sustainable Approach to Unlock University Resources", *D-Lib Magazine, The Magazine of Digital Library Research*, 1996.

[10] Greenstone Digital Library Software, 2007, [Online]. Available: http://www.greenstone.org (accessed 21 June 2010).

[11] Goi S.S. and Zainab A.N., "Postgraduate Research in the Humanities at the University of Malaya", *Malaysian Journal of Library & Information Science*, vol. 2, no. 1, pp. 71-80, 1997.

[12] Hilmi M.R. and Zainab A. N., "Creating a Digital Library to Handle Malay Manuscript Using Greenstone", *Proceedings of the International Conference on Libraries, Information and Society*, pp. 223-231, 2007.

[13] IMR Information Management Research, Alchemy Web Server, 1998 [Online]. Available: http://www.paperless.com/pdf/alchemyweb.pdf (accessed 6 April 2009)

[14] [Lariviere V., Zuccala A. and Archambault E. "The Declining Scientific Impact of Thesis: Implications for Electronic Thesis and Dissertation Repositories and Graduate Studies", *Journal of Scientometrics*, vol. 74, no. 1, pp. 109-121, 2008.

[15] Lee S. and Cho S.Y., "Digital Libraries in Korea", *International Conference on Digital Libraries: Research and Practice*, Kyoto, pp. 130-135, 2001.

[16] Mendes M. J., Suomi R and Passos C., "Digital Communities in a Networked Society: E-Commerce, E-Business, and E-Government: *The Third IFIP Conference on E-Commerce, E-Business, and E-Government*, Saao Paulo, Brazil, 2003.

[17] MOHE (Ministry of Higher Education, Malaysia) Statistic on Local University Student Intake, 2008, [Online]. Available: http://www.mohe.gov.my/web_statistik/statistik_pdf_2008_05/ipta_2-4.pdf (accessed 17 October 2008).

[18] Moore R. W., Wan, M. and Rajasekar A., "Storage Resource Broker; Generic Software Infrastructure for Managing Globally Distributed Data," *Local to Global Data Interoperability - Challenges and Technologies*, pp. 65-69, 2005.

[19] MyTO (Malaysian Theses Online) Official Website, 2009, [Online]. Available: http://www.perpun.net.my/myto (accessed 3 February 2009)

[20] NDLTD Research, *Networked Digital Library of Theses and Dissertations – Research Project*, 2008, [Online]. Available: http://www.ndltd.org/community/research (accessed 20 October 2008)

[21] Rajasekar A., Wan M. and Moore R., "MySRB and SRB - Components of a Data Grid," *Proceedings of 11th IEEE International Symposium High Performance Distributed Computing*, pp. 301-310, 2002.

[22] Rajasekar A., Wan M. and Moore R., "Storage Resource Broker – Managing Distributed Data in a Grid", *Computer Society of India Journal*, vol. 33, no. 4, pp. 42-54, 2003.

[23] Suleman H. and Fox E.A., "Towards Universal Accessibility of ETDs: Building the NDLTD

Union Archive", *Proceedings of Fifth International Symposium on Electronic Theses and Dissertations,* 2002.

[24] Tanskey R., Bass M. and Stuve D. "The DSpace Institutional Digital Repository System: Current Functionality", *Proceedings of the 2003 Joint Conference on Digital Libraries*, pp. 87-97, 2003

[25] UiTM (University of Technology MARA), *Digital Collections 2007*. [Online]. Available: http://digital.ptar.uitm.edu.my/ (accessed 3 February 2009).

[26] UniMAP (University Malaysia Perlis) Library Digital Repository, 2007. [Online]. Available: http://dspace.unimap.edu.my/ (accessed 3 February 2009).

[27] Wan M., Rajasekar A. and Moore R. "A Simple Mass Storage System for the SRB Data Grid," *Proceedings of the 20th IEEE/11th NASA Goddard Conference on Mass Storage Systems and Technologies*, p. 20-35, 2003.

[28] Webster J., "Theses and Dissertations for the Next Millenium," International Association of Aquatic and Marine Science Libraries and Information Centers, pp. 75-86., 1999.

[29] Witten I. H., Banbridge D. and Tansley R., "StoneD: A Bridge between Greenstone and DSpace", D-Lib Magazine, vol. 11, no. 9, 2005.

Delina Beh Mei Yin received her Bachelor of Information Technology at University of Malaya, Malaysia. She is currently pursuing her Master of Computer Science at University of Malaya, Malaysia.

Maizatul Akmar Ismail received her Bachelor of Information Technology at University of Malaya, and Master of Science in Information System at University Putra Malaysia. She is a senior lecturer in the Faculty of Computer Science and Information Technology at University of Malaya since 2008. Her current research interest include semantic web in education and data grid.

Lip Yee Por received his B. Comp. Sc. and M. Comp. Sc. at University of Malaya, Malaysia. He is a lecturer in the Faculty of Computer Science and Information Technology at University of Malaya since 2004. His research areas include steganography, graphical authentication and grid computing. He is a Member of IEEE since 2007. His biography has been included in Marquis Who's Who in the World.

Ong Sim Ying received her Bachelor of Computer Science in Software Engineering at University of Malaya, Malaysia. She is currently pursuing her PhD at University of Malaya, Malaysia.